Europäisches Patentamt | European Patent Office | Office européen des brevets

Bescheinigung | Certificate | Attestation

Die angehefteten Unterlagen stimmen mit der ursprünglich eingereichten Fassung der auf dem nächsten Blatt bezeichneten europäischen Patentanmeldung überein.

The attached documents are exact copies of the European patent application described on the following page, as originally filed.

Les documents fixés à cette attestation sont conformes à la version initialement déposée de la demande de brevet européen spécifiée à la page suivante.

Patentanmeldung Nr. Patent application No. Demande de brevet n°

02405854.7

Der Präsident des Europäischen Patentamts;
Im Auftrag

For the President of the European Patent Office

Le Président de l'Office européen des brevets
p.o.

R C van Dijk

Anmeldung Nr:
Application no.: 02405854.7
Demande no:

Anmeldetag:
Date of filing: 03.10.02
Date de dépôt:

Anmelder/Applicant(s)/Demandeur(s):

Ecole Polytechnique Fédérale de Lausanne

1015 Lausanne
SUISSE

Bezeichnung der Erfindung/Title of the invention/Titre de l'invention:
(Falls die Bezeichnung der Erfindung nicht angegeben ist, siehe Beschreibung.
If no title is shown please refer to the description.
Si aucun titre n'est indiqué se referer à la description.)

Substrates for O6-alkylguanine-DNA alkyltransferase

In Anspruch genommene Priorität(en) / Priority(ies) claimed /Priorité(s) revendiquée(s)
Staat/Tag/Aktenzeichen/State/Date/File no./Pays/Date/Numéro de dépôt:

Internationale Patentklassifikation/International Patent Classification/
Classification internationale des brevets:

C12Q1/00

Am Anmeldetag benannte Vertragstaaten/Contracting states designated at date of filing/Etats contractants désignées lors du dépôt:

AT BE BG CH CY CZ DE DK EE ES FI FR GB GR IE IT LI LU MC NL PT SE SK TR

Substrates for O^6-Alkylguanine-DNA Alkyltransferase

Field of the Invention

The present invention relates to methods of transferring a label from novel substrates to O^6-alkylguanine-DNA alkyltransferases (AGT) and O^6-alkylguanine-DNA alkyltransferase fusion proteins, and to novel substrates suitable in such methods.

Background of the invention

The mutagenic and carcinogenic effects of electrophiles such as N-methyl-N-nitrosourea are mainly due to the O^6-alkylation of guanine in DNA. To protect themselves against DNA-alkylation, mammals and bacteria possess a protein, O^6-alkylguanine-DNA alkyltransferases (AGT) which repairs these lesions. AGT transfers the alkyl group from the position O-6 of alkylated guanine and guanine derivatives to the mercapto group of one of its own cysteines, resulting in an irreversibly alkylated AGT. The underlying mechanism is a nucleophilic reaction of the S_N2 type which explains why not only methyl groups, but also benzylic groups are easily transferred. As overexpression of AGT in tumour cells is the main reason for resistance to alkylating drugs such as procarbazine, dacarbazine, temozolomide and bis-2-chloroethyl-N-nitrosourea, inhibitors of AGT have been proposed for use as sensitisers in chemotherapy (Pegg *et al.*, Prog Nucleic Acid Res Mol Biol 51: 167-223, 1995).

DE 199 03 895 discloses an assay for measuring levels of AGT which relies on the reaction between biotinylated O^6-alkylguanine derivatives and AGT which leads to biotinylation of the AGT. This in turn allows the separation of the AGT on a streptavidin coated plate and its detection, e.g. in an ELISA assay. The assay is suggested for monitoring the level of AGT in tumour tissue and for use in screening for AGT inhibitors.

Damoiseaux *et al.*, ChemBiochem. 4: 285-287, 2001, disclose modified O^6-alkylated guanine derivatives incorporated into oligodeoxyribonucleotides for use as chemical probes for labelling AGT, again to facilitate detecting the levels of this enzyme in cancer cells to aid in research and in chemotherapy.

PCT/GB02/01636 discloses a method for detecting and/or manipulating a protein of interest wherein the protein is fused to AGT and the AGT fusion protein contacted with an

AGT substrate carrying a label, and the AGT fusion protein detected and optionally further manipulated using the label. Several AGT fusion proteins to be used, general structural principles of the AGT substrate and a broad variety of labels and methods to detect the label useful in the method are described.

Summary of the invention

The invention relates to a method for detecting and/or manipulating a protein of interest, wherein the protein of interest is incorporated into a AGT fusion protein, the AGT fusion protein is contacted with particular AGT substrates carrying a label, and the AGT fusion protein is detected and optionally further manipulated using the label in a system designed for recognising and/or handling the label.

The particular AGT substrates used in the method of the invention are O^6-substituted guanine derivatives or related nitrogen containing hydroxy-heterocycles and their sulfur analogs wherein the O^6-substitutent is an activated methyl derivative suitable for transfer from guanine or the corresponding heterocycle to AGT, and further carrying a label. Activated methyl derivatives are e.g. arylmethyl derivatives suitably substituted in the aryl ring, heteroarylmethyl derivatives suitably substituted in the heteroaryl ring, and allyl type derivatives suitably substituted at the double bond. Suitable substituents of the aryl ring, heteroaryl ring or allylic double bond are linkers connecting a label to the aryl ring, heteroaryl ring or allyl group, preferably linkers which may undergo further modification or cleavage, and also linkers which give rise to dimeric or cyclised AGT substrates. The invention relates also to the novel AGT substrates as such, to methods of manufacture of such novel substrates, and to intermediates useful in the synthesis of such novel AGT substrates.

Detailed description of the invention

In the present invention a protein or peptide of interest is fused to an O^6-alkylguanine-DNA alkyltransferase (AGT). The protein or peptide of interest may be of any length and both with and without secondary, tertiary or quaternary structure, and preferably consists of at least twelve amino acids and up to 2000 amino acids. Examples of such protein or peptide of interest are provided below, and are e.g. enzymes, DNA-binding proteins, transcription regulating proteins, membrane proteins, nuclear receptor proteins, nuclear localization signal proteins, protein cofactors, small monomeric GTPases, ATP-binding

cassette proteins, intracellular structural proteins, proteins with sequences responsible for targeting proteins to particular cellular compartments, proteins generally used as labels or affinity tags, and domains or subdomains of the aforementioned proteins. The protein or peptide of interest is preferably fused to AGT by way of a linker which may be cleaved by an enzyme, e.g. at the DNA stage by suitable restriction enzymes, e.g. AGATCT cleavable by *Bgl II*, and/or linkers cleavable by suitable enzymes at the protein stage, e.g. tobacco etch virus NIa (TEV) protease. Fusion proteins may be expressed in prokaryotic hosts, preferably *E. coli*, or eukaryotic host, e.g. yeast or mammalian cells.

The O^6-alkylguanine-DNA alkyltransferase (AGT) has the property of transferring a label present on a substrate to one of the cysteine residues of the AGT forming part of a fusion protein. In preferred embodiments, the AGT is a known human O^6-alkylguanine-DNA alkyltransferase, hAGT. Murine or rat forms of the enzyme are also considered provided they have similar properties in reacting with a substrate like human AGT. In the present invention, O^6-alkylguanine-DNA alkyltransferase also includes variants of a wild-type AGT which may differ by virtue of one or more amino acid substitutions, deletions or additions, but which still retain the property of transferring a label present on a substrate to the AGT part of the fusion protein. AGT variants may be obtained by chemical modification using techniques well known to those skilled in the art. AGT variants may preferably be produced using protein engineering techniques known to the skilled person and/or using molecular evolution to generate and select new O^6-alkylguanine-DNA alkyltransferases. Such techniques are e.g. saturation mutagenesis, error prone PCR to introduce variations anywhere in the sequence, DNA shuffling used after saturation mutagenesis and/or error prone PCR, or family shuffling using genes from several species.

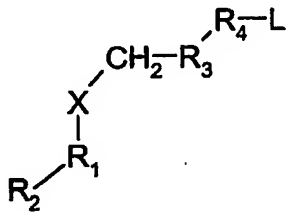
The fusion protein comprising protein of interest and an O^6-alkylguanine-DNA alkyltransferase (AGT) is contacted with a particular substrate having a label. Conditions of reaction are selected such that the AGT reacts with the substrate and transfers the label of the substrate. Usual conditions are a buffer solution at around pH 7 at room temperature, e.g. around 25°C. However, it is understood that AGT reacts also under a variety of other conditions, and those conditions mentioned here are not limiting the scope of the invention.

AGT irreversibly transfers the alkyl group from its substrate, O^6-alkylguanine-DNA, to one of its cysteine residues. A substrate analogue that rapidly reacts with hAGT is O^6-benzylguanine, the second order rate constant being approximately 10^3 sec^{-1} M^{-1}.

Substitutions of O^6-benzylguanine at the C4 of the benzyl ring do not significantly affect the reactivity of hAGT against O^6-benzylguanine derivatives, and this property has been used to transfer a label attached to the C4 of the benzyl ring to AGT.

The label part of the substrate can be chosen by those skilled in the art dependent on the application for which the fusion protein is intended. After contacting the fusion protein comprising AGT with the substrate, the label is covalently bonded to the fusion protein. The labelled AGT fusion protein is then further manipulated and/or detected by virtue of the transferred label.

The particular AGT substrates are compounds of the formula 1



1

wherein R_1-R_2 is a group recognized by AGT as a substrate;
X is oxygen or sulfur;
R_3 is a an aromatic or a heteroaromatic group, or an optionally substituted unsaturated alkyl, cycloalkyl or heterocyclyl group with the double bond connected to CH_2;
R_4 is a linker; and
L is a label, a bond connecting R_4 to R_1 forming a cyclic substrate, or a further group $-R_3-CH_2-X-R_1-R_2$.

In a group R_1-R_2, the residue R_1 is preferably a heteroaromatic group containing 1 to 5 nitrogen atoms, recognized by AGT as a substrate.

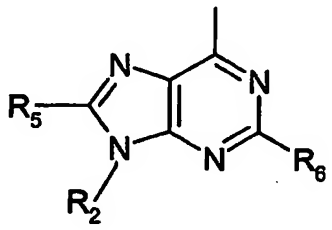
A heteroaromatic group R_1 is mono- or bicyclic and has 5 to 12, preferably 6 or 9 or 10 ring atoms; which in addition to carrying a substituent R_2 may be unsubstituted or substituted by one or more, especially one, two or three further substitutents selected from the group consisting of lower alkyl, such as methyl, lower alkoxy, such as methoxy or ethoxy, hydroxy, oxo, amino, lower alkylamino, di-lower alkylamino, acylamino, halogen, such as chlorine or bromine, halogenated lower alkyl, such as trifluoromethyl, carboxy, lower alkoxycarbonyl, carbamoyl, lower alkylcarbamoyl, or lower alkylcarbonyl.

Lower alkyl is preferably alkyl with from and including 1 up to and including 7, preferably from and including 1 to and including 4, C atoms, and is linear or branched; preferably, lower alkyl is butyl, such as n-butyl, sec-butyl, isobutyl, tert-butyl, propyl, such as n-propyl or isopropyl, ethyl or methyl. Preferably lower alkyl is methyl.

In lower alkoxy, the lower alkyl group is as defined hereinbefore. Lower alkoxy denotes preferably n-butoxy, tert-butoxy, iso-propoxy, ethoxy, or methoxy, in particular methoxy.

Preferably the mono- or bicyclic heteroaromatic group R_1 is selected from 2H-pyrrolyl, pyrrolyl, imidazolyl, benzimidazolyl, pyrazolyl, indazolyl, purinyl, 8-azapurinyl, pyridyl, pyrazinyl, pyrimidinyl, pyridazinyl, 4H-quinolizinyl, isoquinolyl, quinolyl, phthalazinyl, naphthyridinyl, quinoxalyl, quinazolinyl, quinnolinyl, pteridinyl, indolizinyl, 3H-indolyl, indolyl, isoindolyl, triazolyl, tetrazolyl, or benzo[d]pyrazolyl. More preferably the mono- or bicyclic heteroaromatic group R_1 is selected from the group consisting purinyl, 8-azapurinyl, pyridyl, pyrazinyl, pyrimidinyl, and pyridazinyl.

For example the group R_1-R_2 may be a purine radical of the formula 2



2

wherein R_2 is hydrogen, alkyl of 1 to 10 carbon atoms, or a saccharide moiety;
R_5 is hydrogen, halogen, e.g. chloro or bromo, trifluoromethyl, or hydroxy; and
R_6 is hydrogen, hydroxy or unsubstituted or substituted amino.

If R_5 or R_6 is hydroxy, the purine radical is predominantly present in its tautomeric form wherein a nitrogen adjacent to the carbon atom bearing R_5 or R_6 carries a hydrogen atom, the double bond between this nitrogen atom and the carbon atom bearing R_5 or R_6 is a single bond, and R_5 or R_6 is double bonded oxygen, respectively.

A substituted amino group R_6 is lower alkylamino of 1 to 4 carbon atoms or acylamino, wherein the acyl group is lower alkylcarbonyl with 1 to 5 carbon atoms, e.g. acetyl,

propionyl, n- or isopropylcarbonyl, or n-, iso- or tert-butylcarbonyl, or arylcarbonyl, e.g. benzoyl.

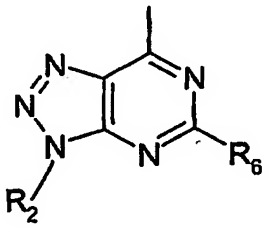
If R_6 is unsubstituted or substituted amino and the residue X connected to the bond of the purine radical is oxygen, the residue of formula 2 is a guanine derivative.

R_2 as alkyl of 1 to 10 carbon atoms is linear or branched and includes lower alkyl of 1 to 4 carbon atoms, e.g. methyl, ethyl, butyl, such as n-butyl, sec-butyl, isobutyl or tert-butyl, and propyl, such as n-propyl or isopropyl. R_2 as alkyl may also be pentyl, hexyl, heptyl, octyl, nonyl, or decyl, e.g. n-hexyl.

A saccharide moiety R_2 is a saccharide monomer or oligomer connected with a spacer of variable length to the N^9 position of the guanine base. The spacer in this context is an alkyl chain preferably from 1 to 15 carbon atoms, a polyethylene glycol spacer consisting of 1 to 200 ethylene glycol units, an amide group –CO-NH-, an ester group –CO-O-, an alkylene group –CH=CH- or a combination of alkyl chain, polyethylene glycol group, amide group, ester group, and/or alkylene group.

In the context of this invention, a saccharide moiety R_2 further includes a β-D-2'-deoxyribosyl, or a β-D-2'-deoxyribosyl being incorporated into a single stranded oligodeoxyribonucleotide having a length of 2 to 99 nucleotides, wherein the guanine derivative R_1 occupies any position within the oligonucleotide sequence.

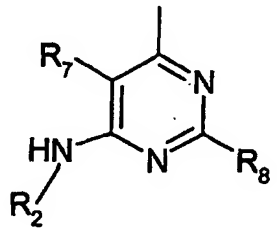
In another preferred embodiment of the invention the group R_1-R_2 is a 8-azapurine radical of the formula 3



3

wherein the substituents R_2 and R_6 have the meaning as defined for R_2 and R_6 under formula 2.

In a further preferred embodiment of the invention the group R_1-R_2 is a pyrimidine radical of the formula 4



4

wherein the substituent R_2 has the meaning as defined under formula 2, and is preferably hydrogen; and

R_7 and R_8 are both independently of one another hydrogen, halogen, e.g. chlorine or bromine, lower alkyl with 1 to 4 carbon atoms, e.g. methyl, amino, or nitro.

X is preferably oxygen.

R_3 as an aromatic or a heteroaromatic group, or an optionally substituted unsaturated alkyl, cycloalkyl or heterocyclyl group is a group sterically and electronically accepted by AGT (in accordance with its reaction mechanism) which allows the covalent transfer of the R_3-R_4-L unit to the fusion protein.

R_3 as an aromatic group is preferably phenyl or naphthyl, in particular phenyl, e.g. phenyl substituted by R_4 in para or meta position.

A heteroaromatic group R_3 is a mono- or bicyclic heteroaryl group comprising zero, one, two, three or four ring nitrogen atoms and zero or one oxygen atom and zero or one sulfur atom, with the proviso that at least one ring carbon atom is replaced by a nitrogen, oxygen or sulfur atom, and which has 5 to 12, preferably 5 or 6 ring atoms; and which in addition to carrying a substituent R_4 may be unsubstituted or substituted by one or more, especially one, further substitutent selected from the group consisting of lower alkyl, such as methyl, lower alkoxy, such as methoxy or ethoxy, halogen, e.g. chlorine, bromine or fluorine, halogenated lower alkyl, such as trifluoromethyl, or hydroxy.

Preferably the mono- or bicyclic heteroaryl group R_3 is selected from 2H-pyrrolyl, pyrrolyl, imidazolyl, benzimidazolyl, pyrazolyl, indazolyl, purinyl, pyridyl, pyrazinyl, pyrimidinyl, pyridazinyl, 4H-quinolizinyl, isoquinolyl, quinolyl, phthalazinyl, naphthyridinyl, quinoxalyl,

quinazolinyl, quinnolinyl, pteridinyl, indolizinyl, 3H-indolyl, indolyl, isoindolyl, oxazolyl, isoxazolyl, thiazolyl, isothiazolyl, triazolyl, tetrazolyl, furazanyl, benzo[d]pyrazolyl, thienyl, and furanyl. More preferably the mono- or bicyclic heteroaryl group is selected from the group consisting of pyrrolyl, imidazolyl, such as 1H-imidazol-1-yl, benzimidazolyl, such as 1-benzimidazolyl, indazolyl, especially 5-indazolyl, pyridyl, e.g. 2-, 3- or 4-pyridyl, pyrimidinyl, especially 2-pyrimidinyl, pyrazinyl, isoquinolinyl, especially 3-isoquinolinyl, quinolinyl, especially 4- or 8-quinolinyl, indolyl, especially 3-indolyl, thiazolyl, triazolyl, tetrazolyl, benzo[d]pyrazolyl, thienyl, and furanyl.

In a particularly preferred embodiment of the invention the heteroaryl group R_3 is triazolyl, especially 1-triazolyl, carrying the further substituent R_4 in the 4- or 5-position, tetrazolyl, especially 1-tetrazolyl, carrying the further substituent R_4 in the 4- or 5-position, or 2-tetrazolyl carrying the further substituent R_4 in 5-position, isoxazolyl, especially 3-isoxazolyl carrying the further substituent R_4 in 5-position, or 5-isoxazolyl, carrying the further substituent R_4 in 3-position, or thienyl, especially 2-thienyl, carrying the further substituent R_4 in 3-, 4- or 5-position, preferably 4-position, or 3-thienyl, carrying the further substituent R_4 in 4-position.

Most preferred is the heteroaryl group R_3 as triazolyl, carrying the substituent R_4 in 4- or 5-position, and also R_3 as 2-thienyl carrying the substituent R_4 in 4- or 5-position.

An optionally substituted unsaturated alkyl group R_3 is 1-alkenyl carrying the further substituent R_4 in 1- or 2-position, preferably in 2-position, or 1-alkynyl. Substituents considered in 1-alkenyl are e.g. lower alkyl, e.g. methyl, lower alkoxy, e.g. methoxy, lower acyloxy, e.g. acetoxy, or halogenyl, e.g. chloro. In a particularly preferred embodiment of the invention R_3 is 1-alkynyl.

An optionally substituted unsaturated cycloalkyl group is a cycloalkyl group with 3 to 7 carbon atoms unsaturated in 1-position, e.g. 1-cyclopentyl or 1-cyclohexyl, carrying the further substituent R_4 in any position. Substituents considered are e.g. lower alkyl, e.g. methyl, lower alkoxy, e.g. methoxy, lower acyloxy, e.g. acetoxy, or halogenyl, e.g. chloro.

An optionally substituted unsaturated heterocyclyl group has 3 to 12 atoms, 1 to 5 heteroatoms selected from nitrogen, oxygen and sulfur, and a double bond in the position connecting the heterocyclyl group to methylene CH_2. Substituents considered are e.g.

lower alkyl, e.g. methyl, lower alkoxy, e.g. methoxy, lower acyloxy, e.g. acetoxy, or halogenyl, e.g. chloro.

In particular, an optionally substituted unsaturated heterocyclyl group is a partially saturated heteroaromatic group as defined hereinbefore for a heteroaromatic group R_3. An example of such a heterocyclyl group is isoxazolidinyl, especially 3-isoxazolidinyl carrying the further substituent in 5-position, or 5-isoxazolidinyl, carrying the further substituent in 3-position.

A linker group R_4 is preferably a flexible linker connecting a label L to the substrate. Linker units are chosen in the context of the envisioned application, i.e. in the transfer of the substrate to a fusion protein comprising AGT. They also increase the solubility of the substrate in the appropriate solvent. The linkers used are chemically stable under the conditions of the actual application. The linker does not interfere with the reaction with AGT nor with the detection of the label L, but may be constructed such as to be cleaved at some point in time after the reaction of the compound of formula 1 with the fusion protein comprising AGT.

A linker R_4 is a straight or branched chain alkylene group with 1 to 300 carbon atoms, wherein optionally

(a) one or more carbon atoms are replaced by oxygen, in particular wherein every third carbon atom is replaced by oxygen, e.g. a poylethyleneoxy group with 1 to 100 ethyleneoxy units;

(b) one or more carbon atoms are replaced by nitrogen carrying a hydrogen atom, and the adjacent carbon atoms are substituted by oxo, representing an amide function –NH-CO-;

(c) one or more carbon atoms are replaced by oxygen, and the adjacent carbon atoms are substituted by oxo, representing an ester function –O-CO-;

(d) the bond between two adjacent carbon atoms is a double or a triple bond, representing a function –CH=CH- or –C≡C-;

(e) one or more carbon atoms are replaced by a phenylene, a saturated or unsaturated cycloalkylene, a saturated or unsaturated bicycloakylene, a bridging heteraromatic or a bridging saturated or unsaturated heterocyclyl group;

(f) two adjacent carbon atoms are replaced by a disulfide linkage -S-S-;

or a combination of two or more, especially two, alkylene and/or modified alkylene groups as defined under (a) to (f) hereinbefore, optionally containing substituents.

Substituents considered are e.g. lower alkyl, e.g. methyl, lower alkoxy, e.g. methoxy, lower acyloxy, e.g. acetoxy, or halogenyl, e.g. chloro.

Further substituents considered are e.g. those obtained when an α-amino acid is incorporated in the linker R_4 wherein carbon atoms are replaced by amide functions –NH-CO- as defined under (b). In such a linker, part of the carbon chain of the alkylene group R_4 is replaced by a group $-(NH\text{-}CHR\text{-}CO)_n-$ wherein n is between 1 and 100 and R represents a varying residue of an α-amino acid.

A further substituent is one which leads to a photocleavable linker R_4, e.g. an o-nitrophenyl group. In particular this substituent o-nitrophenyl is located at a carbon atom adjacent to a amide bond, e.g. in a group $-NH\text{-}CO\text{-}CH_2\text{-}CH(o\text{-}nitrophenyl)\text{-}NH\text{-}CO-$.

A phenylene group replacing carbon atoms as defined under (e) hereinbefore is e.g. 1,2-, 1,3-, or preferably 1,4-phenylene. A saturated or unsaturated cycloalkylene group replacing carbon atoms as defined under (e) hereinbefore is derived from cycloalkyl with 3 to 7 carbon atoms, preferably from cyclopentyl or cyclohexyl, and is e.g. 1,2- or 1,3-cyclopentylene, 1,2-, 1,3-, or preferably 1,4-cyclohexylene, or also 1,4-cyclohexylene being unsaturated e.g. in 1- or in 2-position. A saturated or unsaturated bicycloalkylene group replacing carbon atoms as defined under (e) hereinbefore is derived from bicycloalkyl with 7 or 8 carbon atoms, and is e.g. bicyclo[2.2.1]heptylene or bicyclo[2.2.2]octylene, preferably 1,4-bicyclo[2.2.1]heptylene optionally unsaturated in 2-position or doubly unsaturated in 2- and 5-position, and 1,4-bicyclo[2.2.2]octylene optionally unsaturated in 2-position or doubly unsaturated in 2- and 5-position. A bridging heteroaromatic group replacing carbon atoms as defined under (e) hereinbefore is e.g. triazolidene, preferably 1,4-triazolidene, or isoxazolidene, preferably 3,5-isoxazolidene. A bridging saturated or unsaturated heterocyclyl group replacing carbon atoms as defined under (e) hereinbefore is e.g. derived from an unsaturated heterocyclyl group as defined under R_3 above, e.g. isoxazolidinene, preferably 3,5-isoxazolidinene, or a fully saturated heterocyclyl group with 3 to 12 atoms, 1 to 3 of which are heteroatoms selected from nitrogen, oxygen and sulfur, e.g. pyrrolidinediyl, piperidinediyl, tetrahydrofuranediyl, dioxanediyl, morpholinediyl or terahydrothiophenediyl, preferably 2,5- tetrahydrofuranediyl or 2,5-dioxanediyl.

Cyclic substructures in a linker R_4 reduce the molecular flexibility as measured by the number of rotatable bonds within R_4, which leads to a better membrane permeation rate, important for all *in vivo* labeling applications.

A linker R_4 is preferably a straight chain alkylene group with 1 to 25 carbon atoms or a straight chain polyethylene glycol group with 4 to 100 ethyleneoxy units, optionally attached to the group R_3 by a –CH=CH- or –C≡C- group. Further preferred is a straight chain alkylene group with -1 to 25 carbon atoms wherein carbon atoms are optionally replaced by an amide function –NH-CO-, and carrying a photocleavable subunit, e.g. o-nitrophenyl.

The label part L of the substrate can be chosen by those skilled in the art dependent on the application for which the fusion protein is intended. Labels may be e.g. such that the labelled fusion protein is easily detected or separated from its environment. Other labels considered are those which are capable of sensing and inducing changes in the environment of the labelled fusion protein and/or labels which aid in manipulating the fusion protein by the physical and/or chemical properties specifically introduced by the label to the fusion protein.

Examples of labels L include a spectroscopic probe such as a fluorophore, a chromophore, a magnetic probe or a contrast reagent; a radioactively labelled molecule; a molecule which is one part of a specific binding pair which is capable of specifically binding to a partner; a molecule that is suspected to interact with other biomolecules; a library of molecules that are suspected to interact with other biomolecules; a molecule which is capable of crosslinking to other molecules; a molecule which is capable of generating hydroxyl radicals upon exposure to H_2O_2 and ascorbate, such as a tethered metal-chelate; a molecule which is capable of generating reactive radicals upon irradiation with light, such as malachite green; a molecule covalently attached to a solid support, where the support may be a glass slide, a microtiter plate or any polymer known to those proficient in the art; a nucleic acid or a derivative thereof capable of undergoing base-pairing with its complementary strand; a lipid or other hydrophobic molecule with membrane-inserting properties; a biomolecule with desirable enzymatic, chemical or physical properties; or a molecule possessing a combination of any of the properties listed above.

When the label L is a fluorophore, a chromophore, a magnetic label, a radioactive label or the like, detection is by standard means adapted to the label and whether the method is used *in vitro* or *in vivo*. The method can be compared to the applications of the green fluorescent protein (GFP) which is genetically fused to a protein of interest and allows protein investigation in the living cell. Particular examples of labels L are also boron compounds displaying non-linear optical properties, or a member of a FRET pair which changes its spectroscopic properties on reaction of the labelled substrate with the AGT fusion protein.

Depending on the properties of the label L, the fusion protein comprising protein of interest and AGT may be bound to a solid support. The label of the substrate reacting with the fusion protein comprising AGT may already be attached to a solid support when entering into reaction with AGT, or may subsequently, i.e. after transfer to AGT, be used to attach the AGT fusion protein to a solid support. The label may be one member of a specific binding pair, the other member of which is attached or attachable to the solid support, either covalently or by any other means. A specific binding pair considered is e.g. biotin and avidin or streptavidin. Either member of the binding pair may be the label L of the substrate, the other being attached to the solid support. Further examples of labels allowing convenient binding to a solid support are e.g. maltose binding protein, glycoproteins, FLAG tags, or reactive substituents allowing chemoselective reaction between such substituent with a complementary functional group on the surface of the solid support. Examples of such pairs of reactive substituents and complementary functional group are e.g. amine and activated carboxy group forming an amide, azide and a propiolic acid derivative undergoing a 1,3-dipolar cycloaddition reaction, amine and another amine functional group reacting with an added bifunctional linker reagent of the type of activated bis-dicarboxylic acid derivative giving rise to two amide bonds, or other combinations known in the art.

Examples of a convenient solid support are e.g. glass surfaces such as glass slides, microtiter plates, and suitable sensor elements, in particular functionalized polymers (e.g. in the form of beads), chemically modified oxidic surfaces, e.g. silicon dioxide, tantalum pentoxide or titanium dioxide, or also chemically modified metal surfaces, e.g. noble metal surfaces such as gold or silver surfaces. Irreversibly attaching and/or spotting AGT substrates may then be used to attach AGT fusion proteins in a spatially resolved manner, particularly through spotting, on the solid support representing protein microarrays, DNA microarrays or arrays of small molecules.

When the label L is capable of generating reactive radicals, such as hydroxyl radicals, upon exposure to an external stimulus, the generated radicals can then inactivate the AGT fusion proteins as well as those proteins that are in close proximity of the AGT fusion protein, allowing to study the role of these proteins. Examples of such labels are tethered metal-chelate complexes that produce hydroxyl radicals upon exposure to H_2O_2 and ascorbate, and chromophores such as malachite green that produce hydroxyl radicals upon laser irradiation. The use of chromophores and lasers to generate hydroxyl radicals is also known in the art as chromophore assisted laser induced inactivation (CALI). In the present invention, labelling AGT fusion proteins with chromophores such as malachite green and subsequent laser irradiation inactivates the AGT fusion protein as well as those proteins that interact with the AGT fusion protein in a time-controlled and spatially-resolved manner. This method can be applied both *in vivo* or *in vitro*. Furthermore, proteins which are in close proximity of the AGT fusion protein can be identified as such by either detecting fragments of that protein by a specific antibody, by the disappearance of those proteins on a high-resolution 2D-electrophoresis gels or by identification of the cleaved protein fragments via separation and sequencing techniques such as mass spectrometry or protein sequencing by N-terminal degradation.

When the label L is a molecule that can cross-link to other proteins, e.g. a molecule containing functional groups such as maleimides, active esters or azides and others known to those proficient in the art, contacting such labelled AGT substrates with AGT fusion proteins that interact with other proteins (*in vivo* or *in vitro*) leads to the covalent cross-linking of the AGT fusion protein with its interacting protein via the label. This allows the identification of the protein interacting with the AGT fusion protein. Labels L for photo cross-linking are e.g. benzophenones. In a special aspect of cross-linking the label L is a molecule which is itself an AGT substrate leading to dimerization of the AGT fusion protein. The chemical structure of such dimers may be either symmetrical (homodimers) or unsymmetrical (heterodimers).

Other labels L considered are for example fullerenes, boranes for neutron capture treatment, nucleotides or oligonucleotides, e.g. for self-adressing chips, peptide nucleic acids, and metal chelates, e.g. platinum chelates that bind specifically to DNA.

The present invention provides a method to label AGT fusion proteins both *in vivo* as well as *in vitro*. The term *in vivo* labelling of a AGT fusion protein includes labelling in all

compartments of a cell as well as of AGT fusion proteins pointing to the extracellular space. If the labelling of the AGT fusion protein is done *in vivo* and the protein fused to the AGT is a membrane protein, more specifically a plasma membrane protein, the AGT part of the fusion protein can be attached to either side of the membrane, e.g. attached to the cytoplasmic or the extracellular side of the plasma membrane.

If the labelling is done *in vitro*, the labelling of the fusion protein can be either performed in cell extracts or with purified or enriched forms of the AGT fusion protein.

If the labelling is done *in vivo* or in cell extracts, the labelling of the endogenous AGT of the host is advantageously taken into account. If the endogenous AGT of the host does not accept O^6-alkylguanine derivatives or related compounds as a substrate, the labelling of the fusion protein is specific. In mammalian cells, e.g. in human, murine, or rat cells, labelling of endogenous AGT is possible. In those experiments where the simultaneous labelling of the endogenous AGT as well as of the AGT fusion protein poses a problem, known AGT-deficient cell lines can be used.

In a particular aspect, the present invention provides a method of determining the interaction of a candidate compound or library of candidate compounds and a target protein or library of target proteins. Examples of candidate compounds and target proteins include ligands and proteins, drugs and targets of the drug, or small molecules and proteins. In this particular method of the invention, the protein of interest fused to the AGT comprises a DNA binding domain of a transcription factor or an activation domain of a transcription factor. The putative protein target of the substances or library of proteins is linked to either of the DNA binding domain or the activation domain of the transcription factor in a way a functional transcription factor can be formed, and the label L of the AGT substrate according to the invention is a candidate compound or library of candidate compounds suspected of interacting with the target substance or substances. The candidate compound or library of candidate compounds being part of the substrate is then transferred to the AGT fusion protein. On transfer the AGT fusion protein(s) comprising the target substance(s) now are labelled with the candidate compound(s). The interaction of a candidate compound joined to the AGT fusion protein with the target protein fused to either the DNA binding domain or the activation domain leads to the formation of a functional transcription factor. The activated transcription factor can then drive the expression of a reporter which, if the method is carried out in cells, can be detected if the expression of the reporter confers a selective advantage on the cells. In particular

embodiments, the method may involve one or more further steps such as detecting, isolating, identifying or characterising the candidate compound(s) or target substance(s).

In a specific example the label L is a drug or a biological active small molecule that binds to an yet unidentified protein Y. A cDNA library of the organism which is expected to express the unknown target protein Y is fused to the activation domain of a transcription factor, and the AGT is fused to the DNA binding domain of a transcription factor. Adding the AGT substrate of the invention comprising such a label L leads to the formation of a functional transcription factor and gene expression only in the case where this molecule binds to its target protein Y present in the cDNA library and fused to the activation domain. If gene expression is coupled to a selective advantage, the corresponding host carrying the plasmid with the gene coding for the target protein Y of the drug or bioactive molecule can be identified.

In a further specific example the label L is a library of chemical molecules. The library is expected to contain yet unidentified compounds that bind to a known drug target protein Y under *in vivo* conditions. The target protein Y is fused to the activation domain of a transcription factor and the AGT is fused to the DNA binding domain of a transcription factor. Adding the substrate carrying the library of chemical compounds leads to the formation of a functional transcription factor and gene expression only in the case where the label (i.e. a compound in the chemical library) binds to its target protein Y fused to the activation domain. If gene expression is coupled to a selective advantage, those molecules of the library leading to the growth of the host can be identified.

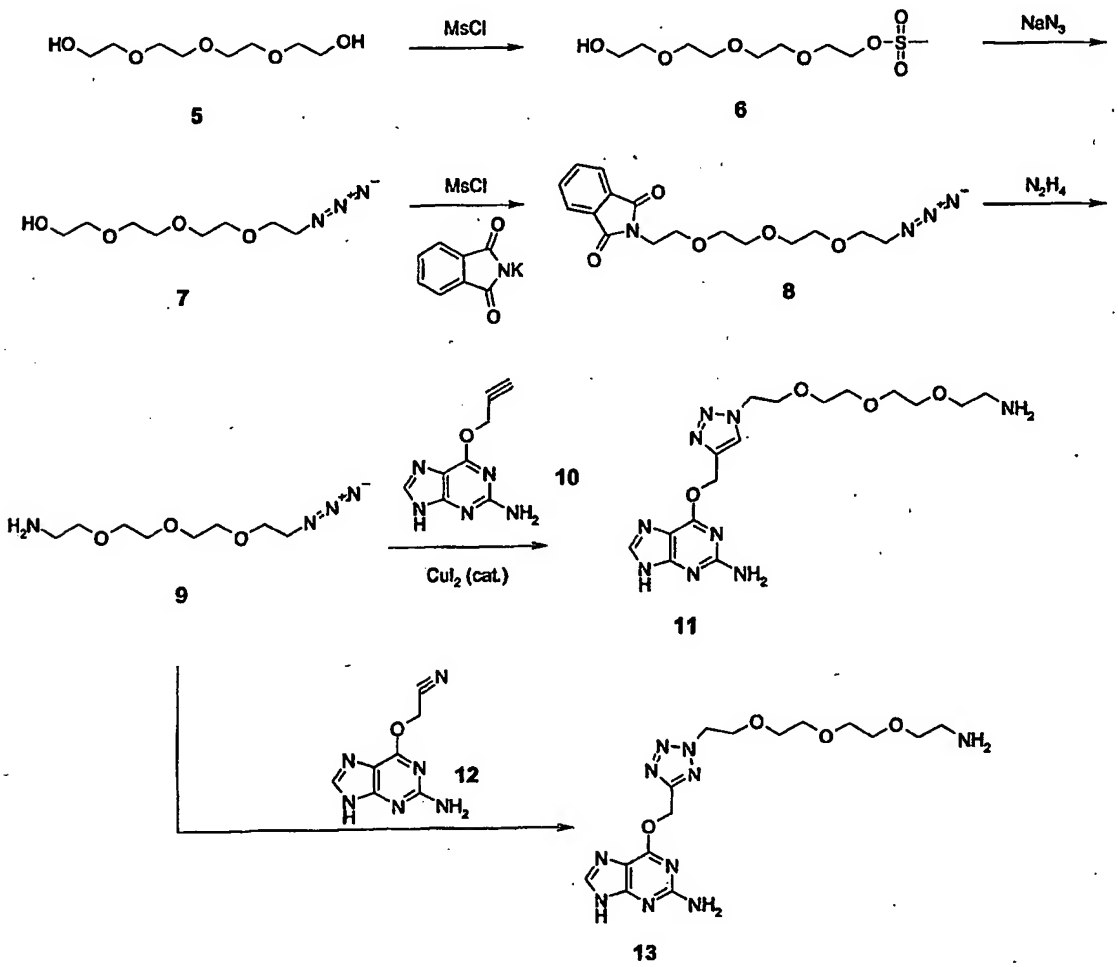
In the case where L is a bond connecting R_4 to R_1 forming a cyclic substrate, a preferred compound is the cyclic substrate wherein the bond from R_4 to R_1 is a bond connecting the linker R_4 to an amino group R_6 as defined under formula 2. In such a preferred cyclic substrate, R_2 is preferably an oligonucleotide, i.e. a β-D-2'-deoxyribosyl being incorporated into a single stranded oligodeoxyribonucleotide having a length of 2 to 99 nucleotides as detailed above. This oligonucleotide may be further chemically modified so that it can be detected and functions therefore as a label. The chemical modification of substituents might be of the same nature as mentioned above for the label L.

In the case where L is a further group $-R_3-CH_2-X-R_1-R_2$, the substrate is a dimeric compound leading to a dimerised fusion protein on reaction with a fusion protein comprising AGT. In the subunit L as a residue $-R_3-CH_2-X-R_1-R_2$, the meaning of R_1, R_2,

R_3 and X may be identical with the corresponding meaning in the other group R_2-R_1-X-CH_2-R_3-, representing a homodimer, or different, representing a heterodimer.

Methods of manufacture of novel substrates are also an object of this invention.

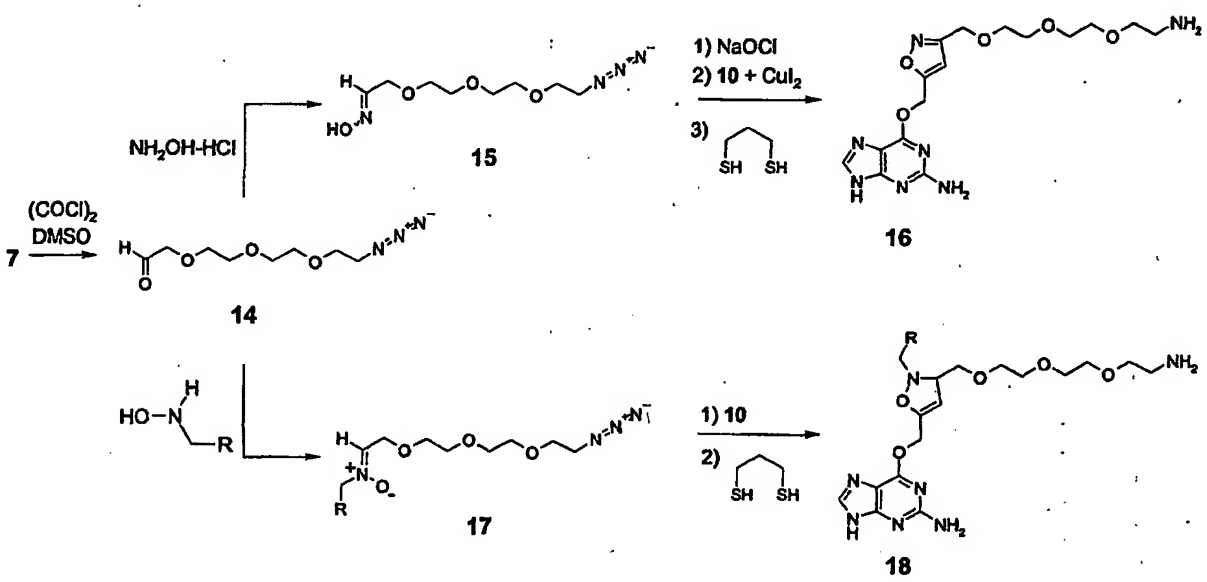
The synthesis of an intermediate useful in the synthesis of compounds of formula 1 wherein R_3 is a tetrazolyl group, an isoxazolyl group or an isoxazolidinyl group is summarized in Scheme 1 and 2.



Scheme 1

The azido compound **7** is prepared from commercially available tetraethylene glycol **5** by mesylation (methanesulfonyl chloride, Et_3N) followed by reaction with sodium azide in ethanol. **7** is again mesylated and subjected to a Gabriel amine synthesis to give azido-

amine **9** (Carolay *et al.*, J. Org. Chem. 56: 4326-4329, 1991). The Cu(I)-catalyzed 1,3-dipolar cycloaddition between azide **9** and the acetylene derivative **10** (Griffin *et al.*, J. Med. Chem. 43: 4071-4083, 2000) yields the 1,4-substituted triazole **11**. Alternatively the azide **9** and the cyano derivative **12** react under Lewis acid catalysis ($ZnBr_2$) to form tetrazole **13**.

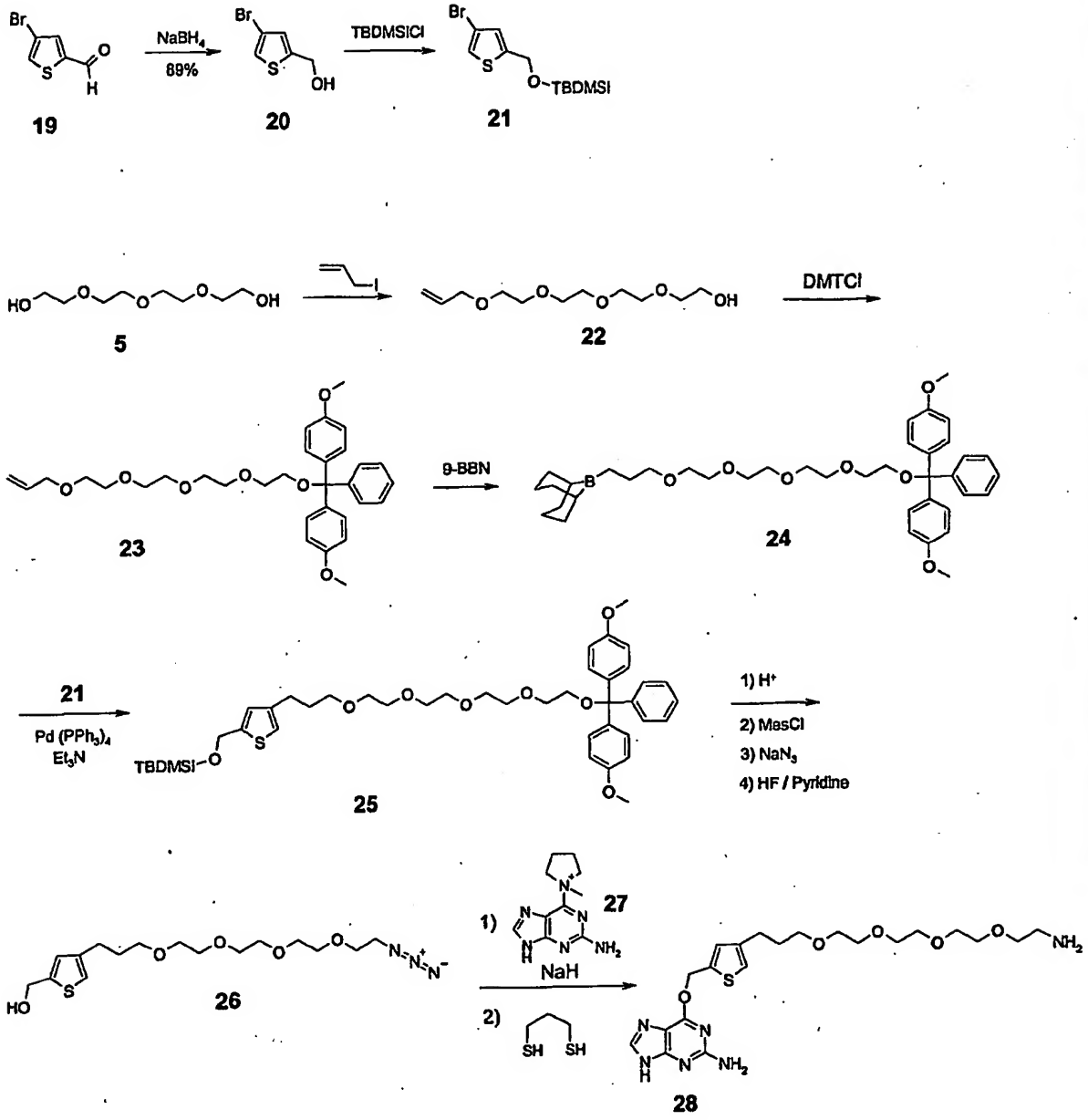


Scheme 2

Azide **7** is transformed to the central building block, the aldehyde **14**, by means of a Swerns oxidation (oxalylchloride, DMSO, Et_3N). The reaction of **14** with a hydroxylamine derivative yields the nitrone **17**, which upon reaction with the acetylene derivative **10** forms the class of isoxazolidines **18**.

From aldehyde **14**, the oxime is formed as an equimolar mixture of isomers. The corresponding nitrile-oxide is formed *in situ* by oxidation with sodium hypochlorite followed by reaction with **10** to yield the isoxazole **16**.

The synthesis of an intermediate useful in the synthesis of compounds of formula 1 wherein R_3 is a thienyl group is summarized in Scheme 3.

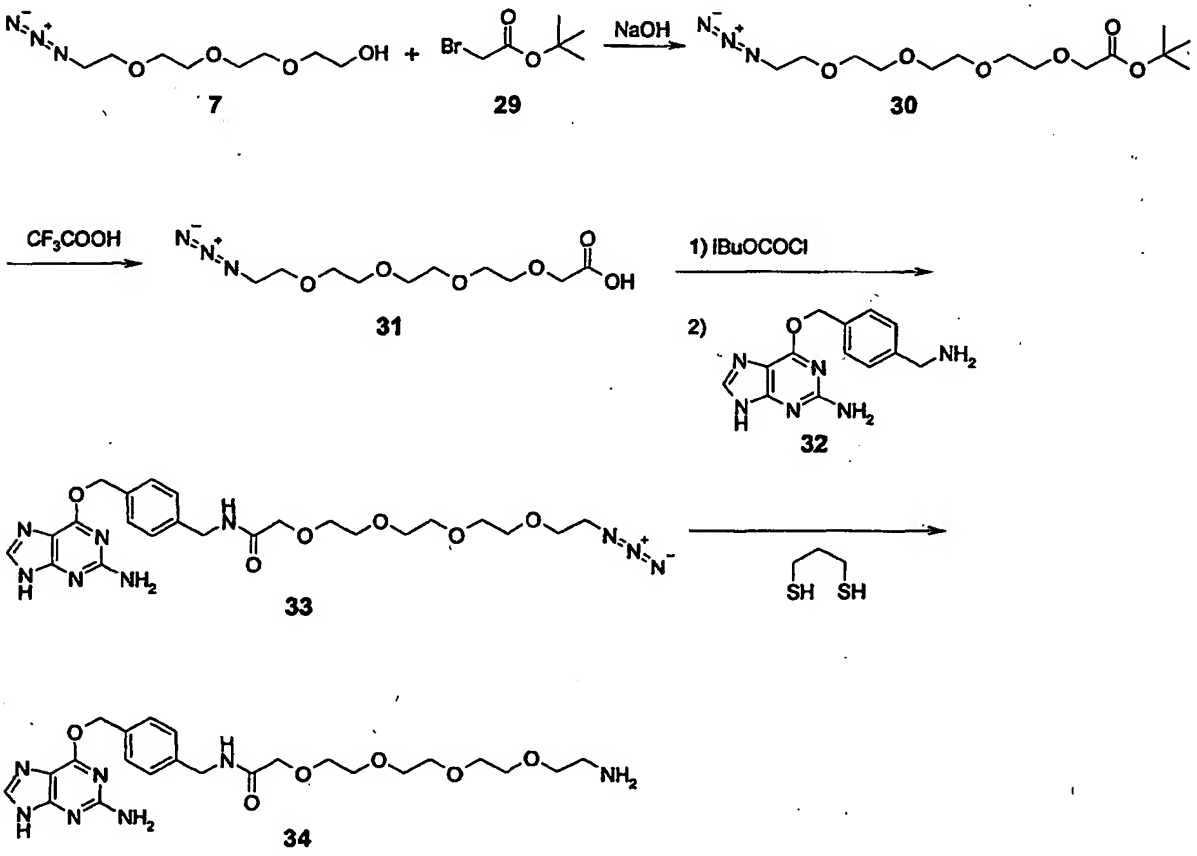


Scheme 3

The commercially available tetraethylene glycol **5** is monofunctionalized through the reaction with one equivalent of allyl iodide under strongly basic conditions to yield **22** which is further dimethoxytrityl (DMT)-protected to **23**. This intermediate allows the palladium catalyzed Suzuki coupling with thiophene derivative **21** to the fully protected compound **25**.

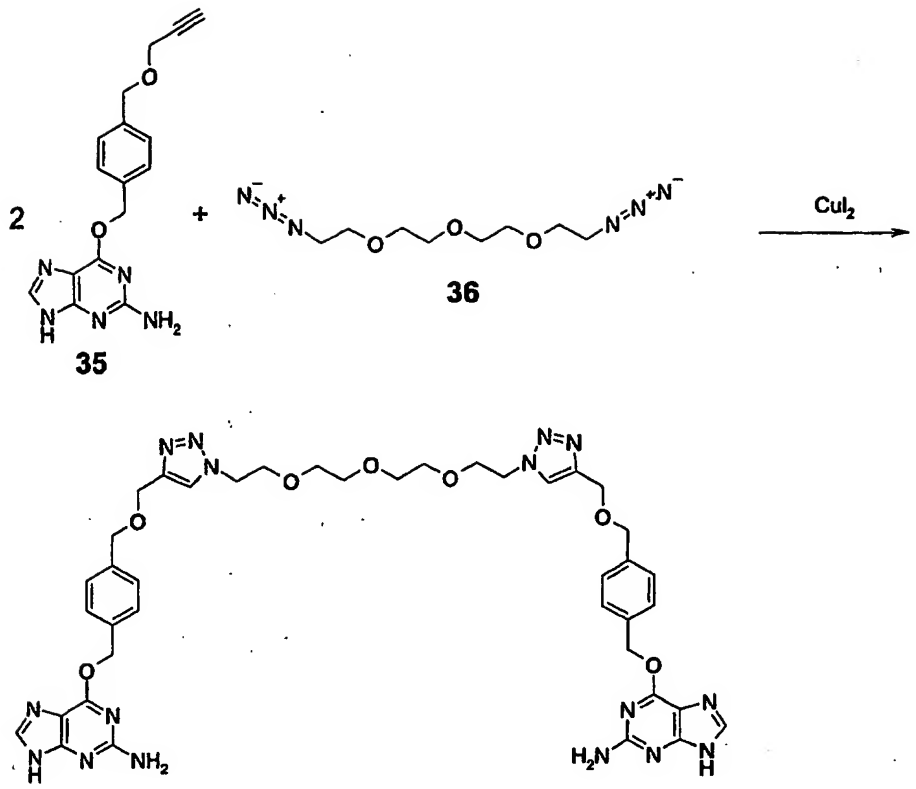
Monodeprotection of the DMT-group and subsequent mesylation (MsCl, Et_3N) followed by the reaction with sodium azide in ethanol gives the protected azide which is deprotected with HF/pyridine to **26**. Coupling of the free hydroxy group with the activated guanine-cation **27** leads to an azido-intermediate which serves as a precursor for different functionalization strategies. Finally reduction of the azide to amine **28** allows the introduction of the label unit L or the coupling to different surfaces.

The synthesis of an intermediate useful in the synthesis of compounds of formula 1 wherein R_3 is a phenylene group is summarized in Scheme 4.



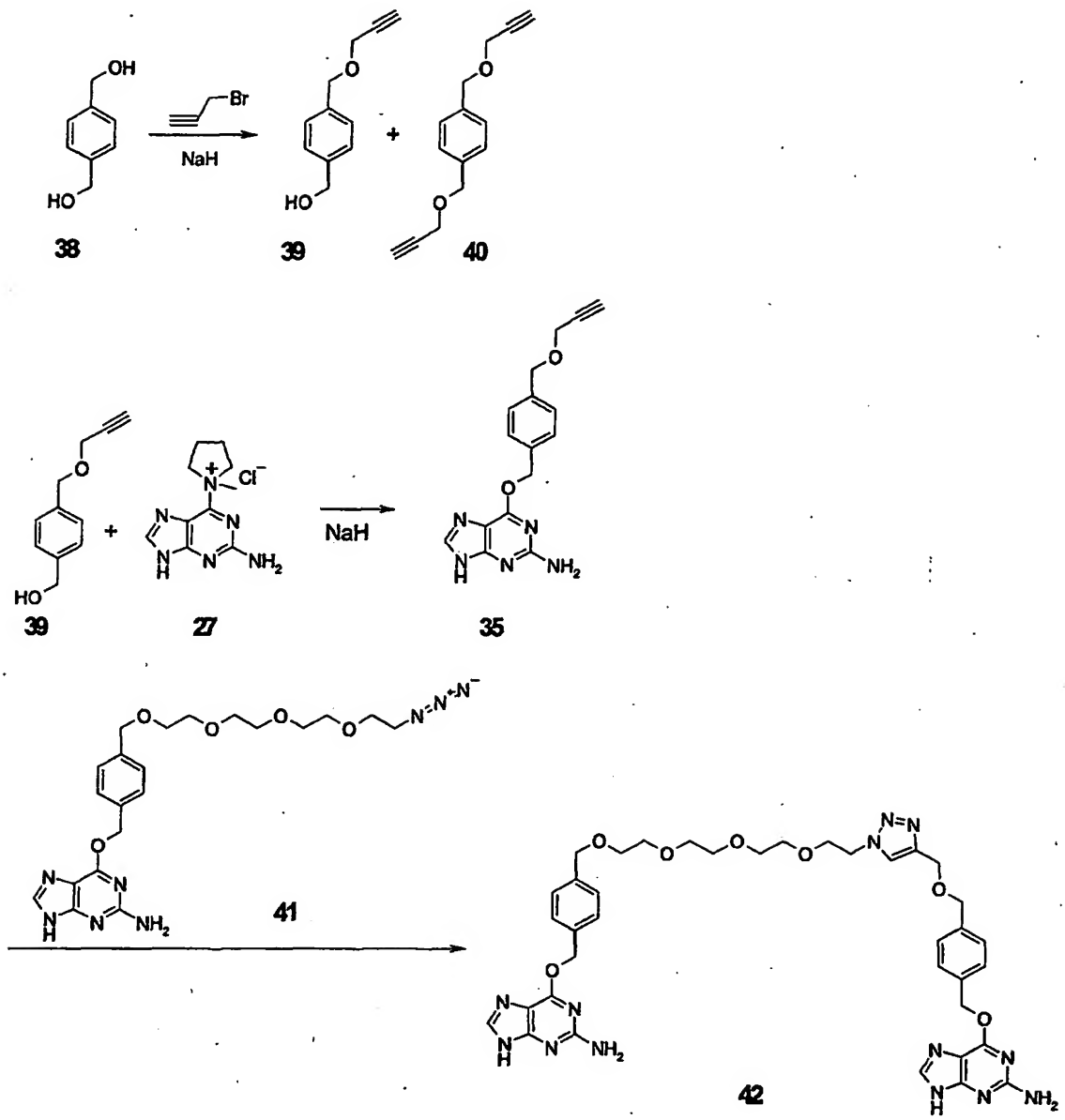
Scheme 4

A compound of formula 1 wherein R_1 is guanine, R_2 is hydrogen, R_3 is triazolyl, R_4 is a triethyleneoxy unit and L is $-R_3-CH_2-X-R_1-R_2$ is prepared as shown in Scheme 5:

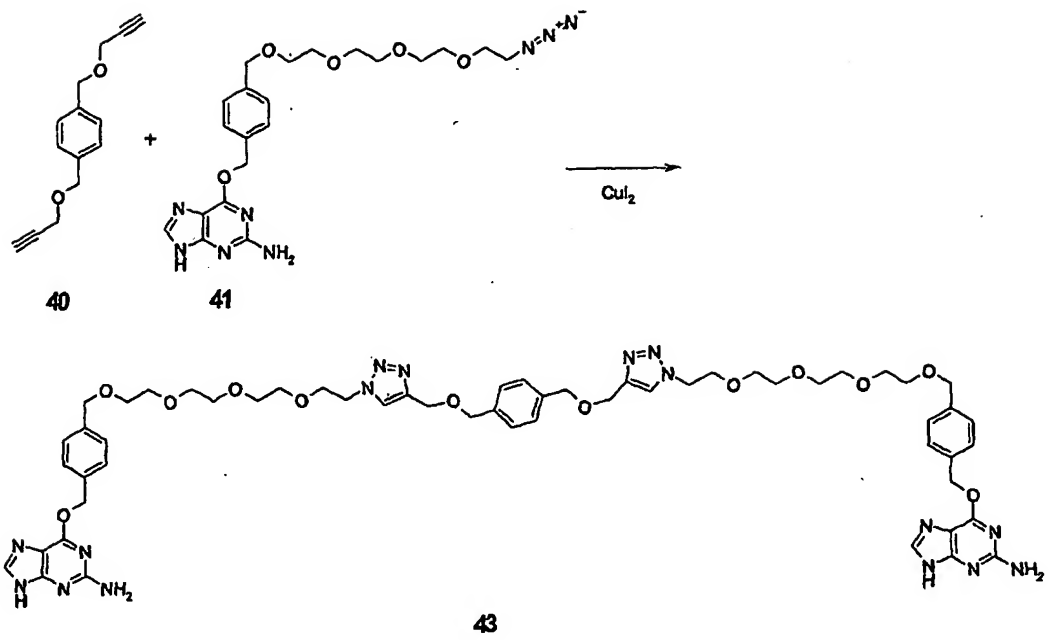


Scheme 5

A compound of formula 1 wherein R_1 is guanine, R_2 is hydrogen, R_3 is 1,4-phenylene, R_4 is a pentaethyleneoxy unit further comprising a triazole group and L is $-R_3-CH_2-X-R_1-R_2$ is prepared as shown in Scheme 6, 7 and 8:



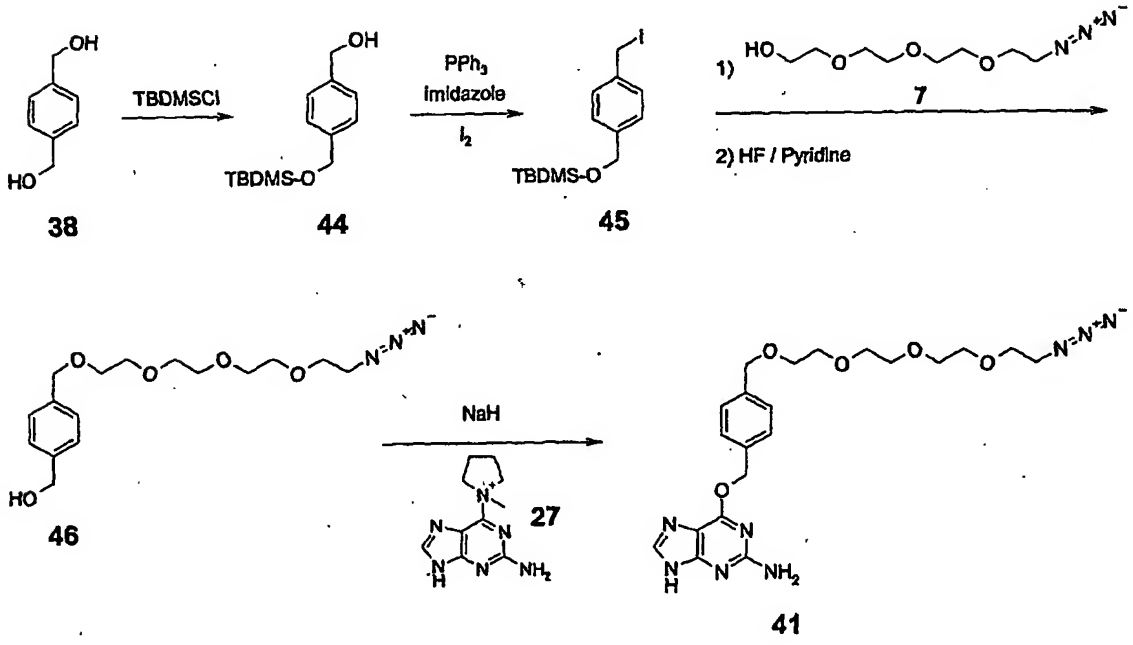
Scheme 6



40 41

43

Scheme 7



38 44 45 46 41

Scheme 8

Examples

Example 1: Preparation of glass slides for the covalent attachment of AGT substrates and subsequent covalent immobilisation of AGT-fusion proteins for the preparation of protein microarrays.

A commercially available microscope glass slide (SiO_2) is cleaned thoroughly with methylene chloride, acetone, H_2O_2/H_2SO_4 in a ultrasonic bath, with bi-distilled water. It is aminosilylated using 3-aminopropyltriethoxysilane in a solvent mixture ethanol/water (95:5) for 1 h following a published procedure, then treated with a solution of disuccinimidyl glutarate (10 mM) in methylene chloride / N-ethyldiisopropylamine (100:1) for 2 h under argon and at room temperature. The surface is washed several times with methylene chloride. The glass surface bearing activated carboxy functions is incubated for 4 h with a solution of an AGT-substrates bearing a free amino group (in methanol, 10 mM, supplemented with triethylamine). The slides are washed at least three times with methanol to yield a surface with the corresponding AGT-substrate covalently attached by an amide bond. To avoid side reactions in further use of the slides, all unreacted succinimidyl groups are quenched by addition of 6-aminohexanol (100 mM in DMF).

Example 2: 1-Azido-11-hydroxy-3,6,9-trioxaundecane (7) and 1,11-Diazido-3,6,9-trioxaundecane (36).

A solution of 50.0 g (260 mM) of tetraethylene glycol and 50 mL triethylamine in 200 mL of dry diethyl ether is cooled to 0° C under an argon atmosphere, and 15.0 g (130 mM) methanesulfonyl chloride is added over a 3 h period and stirred at room temperature for 20 min. The solvent is removed *in vacuo*, and 300 mL 95% ethanol and 18.0 g (280 mM) sodium azide are added. The mixture is heated to reflux for 24 h, cooled to room temperature and concentrated *in vacuo*. The remaining mixture is diluted with 400 mL dichloromethane, washed with brine and dried over $MgSO_4$. After concentration *in vacuo* the crude mixture of mono- and diazide is purified by silica gel chromatography (petrol ether/ethyl acetate 3:1) yielding 15.03 g (68.5 mmol, 26%) monoazide and 3.46 g (14.18 mmol, 5.5%) diazide.

Example 3: 1-Azido-11-phtalimido-3,6,9-trioxaundecane (8)

A solution of 1.17 g (5.35 mmol) 1-azido-11-hydroxy-3,6,9-trioxaundecane (7)and 1.2 mL triethylamine in 35 mL methylene chloride is cooled to 0° C, and 0.5 mL (6.45 mmol) methanesulfonyl chloride is added dropwise *via* a syringe over a 20 min period. The mixture is warmed to room temperature and stirred for 1.5 h. The mixture is then washed

twice with 10 mL of saturated aqueous $NaHCO_3$ and three times with 5 mL of water. The organic layer is dried and concentrated *in vacuo* to yield 1.5 g **8** as a yellow oil which is used without further purification.

Example 4: 4-Bromothenyl alcohol (**20**)

5.0 g (26.17 mmol) 4-bromothiophene-2-carboxaldehyde (**19**) is dissolved in 75 mL iso-propanol, and 1.11 g (29.31 mmol) $NaBH_4$ are added at once and the mixture stirred for 2 h. 20 mL saturated aqueous NH_4Cl is added, the solid removed by filtration and the mixture concentrated *in vacuo*. The product is purified by silica gel chromatography (petrol ether/ethyl acetate 10:1), yielding 4.64 g **20** (24.07 mmol, 92%) as a colorless solid.

Example 5: 15-Hydroxy-4,7,10,13-tetraoxa-1-pentadecene (**22**)

2.3 g (19.5 mmol) potassium tert-butoxide is dissolved in 500 mL dry THF, and 7.18 g (37 mmol) tetraethylene glycol is added dropwise. After stirring for 30 min, a solution of 3.31 g (19.7 mmol) allyl iodide in 60 mL dry THF is added over 1 h, and stirring is continued for 24 h. The crude mixture is filtered over silica gel and the solvent removed *in vacuo*. The product is purified by silica gel chromatography (gradient: petrol ether/ethyl acetate 10:1 → ethyl acetate), yielding 2.41 g **22** (10.3 mmol, 27%) as a colorless liquid.

Example 6: 1-(2-Amino-7*H*-purin-6-yl)-1-methyl-pyrrolidinium chloride (**27**)

1.0 g (5.9 mmol) 6-chloroguanine is dissolved in 40 mL DMF at 40°C. After cooling to room temperature, 1.4 mL 1-methylpyrrolidine (13.2 mmol) are added, and the reaction mixture is stirred for 18 h. 2 mL of acetone are added to complete precipitation. The solid is filtered, washed with ether and dried *in vacuo*, yielding 1.03 g **27** (3.9 mmol, 66 %).

Example 7: O^6-(4-Aminomethyl-benzyl)guanine (**32**)

a) 4-(Aminomethyl)-benzyl alcohol: 2.83 g $LiAlH_4$ (74.5 mmol) are suspended in 150 mL dry ether and 1.9 mL H_2SO_4 (100 %, 37.2 mmol) are added dropwise and under cooling. The mixture is stirred for 1 h at room temperature, followed by dropwise addition of 2.0 g (12.4 mmol) 4-cyanobenzoate in 12 mL ether. After 2 h of refluxing the reaction is quenched with 20 mL water followed by 7.4 g NaOH in 60 mL water. The organic layer is decanted, and the aqueous layer extracted with ether and ethyl acetate. The organic layer is dried over $MgSO_4$, the solvent is removed and the product dried *in vacuo*, yielding 0.92 g (6.7 mmol, 54 %).

b) 2,2,2-Trifluoro-*N*-(4-hydroxymethyl-benzyl)-acetamide: To a solution of 866 mg (6.3 mmol) 4-(aminomethyl)-benzyl alcohol and 880 µL (6.3 mmol) triethylamine in 10 mL dry methanol 980 µL (8.2 mmol) trifluoroacetic acid ethyl ester are added dropwise. The reaction mixture is stirred for 45 min, diluted with 10 mL ethyl acetate and 10 mL water. The aqueous layer is extracted with ethyl acetate and the combined organic layers are washed with saturated NaCl and dried over Na_2SO_4. After removal of the solvents *in vacuo* the crude product is purified by flash column chromatography (ethyl acetate/cyclohexane 1:2). Yield: 1.32 g (5.7 mmol, 90 %).

c) *N*-[4-(2-Amino-7H-purin-6-yloxymethyl)-benzyl]-2,2,2-trifluoroacetamide: 592 mg (2.54 mmol) 2,2,2-trifluoro-*N*-(4-hydroxymethyl-benzyl)acetamide are dissolved in dry DMF under argon atmosphere, and 599 mg (5.33 mmol) potassium tert-butoxide are added. 300 mg (1.18 mmol) 1-(2-amino-7H-purin-6-yl)-1-methylpyrrolidinium chloride (**27**) are then added and the solution stirred for 3 h. After removal of the solvent *in vacuo* the crude product is purified by flash column chromatography (300 mL methanol/dichloromethane 1:50, 500 mL methanol/dichloromethane 1:10). Yield: 382 mg (1.04 mmol, 88 %).

d) O^6-(4-Aminomethyl-benzyl)guanine (**32**): 335 mg (0.91 mmol) N-[4-(2-amino-7H-purin-6-yloxymethyl)-benzyl]-2,2,2-trifluoroacetamide are suspended in 34 mL methanol and 2 mL water. After addition of 656 mg (4.75 mmol) of K_2CO_3 the reaction mixture is refluxed for 2 h. The solvents are removed *in vacuo* and the product is purified by flash column chromatography (methanol/triethylamine/dichloromethane 1:0.05:5). Yield: 209 mg **32** (0.77 mmol, 85 %).

Example 8: O^6-(4-Prop-2-ynyloxymethyl-benzyl)guanine (**35**)

662 mg (3.8 mmol) 4-(prop-2-ynyloxymethyl)-benzyl alcohol (**39**) is dissolved in 3 mL dry DMSO, and 61 mg NaH are added in small portions over 5 min. 300 mg (1.27 mmol) 1-(2-amino-7*H*-purin-6-yl)-1-methylpyrrolidinium chloride (**27**) is added and the mixture is stirred for additional 4 h. The reaction is quenched with 0.2 mL of acetic acid, evaporated to dryness and purified by flash column chromatography (gradient: CH_2Cl_2/MeOH 50:1→10:1) to yield 188 mg **35** (0.61 mmol, 53%).

Example 9: Homo-benzylguanine-dimer **37**

To a solution of 50.0 mg (0.162 mmol) O^6-(4-prop-2-ynyloxymethyl-benzyl)guanine (**35**) and 19.7 mg (0.081 mmol) 1,11-diazido-3,6,9-trioxaundecane (**36**) in 0.5 mL DMF is added a suspension of 15.43 mg (0.081 mmol) CuI in 0.15 mL of water. The mixture is stirred at room temperature for 24 h.

Example 10: 4-(Prop-2-ynyloxymethyl)benzyl alcohol (**39**) and 1,4-bis-(prop-2-ynyloxymethyl)benzene (**40**)

To a solution of 2.5 g (18.1 mmol) 4-hydroxymethylbenzyl alcohol is added 477.5 mg (19.9 mmol) NaH in small portions over 20 min. 2.15 mL of a propargyl bromide solution (80% in toluene) is added dropwise and stirred for 15 h. 100 mL of water are added to the mixture, and the products extracted with diethyl ether. The combined phase is dried and the solvent removed *in vacuo*. The separation of the products is achieved by silica gel chromatography (petrol ether/ethyl acetate 4:1) yielding 1.08 g **39** (6.17 mmol, 34%) and 1.05 g **40** (4.94 mmol, 27%).

Example 11: 4-[(tert-Butyldimethylsilyloxy)methyl]benzyl alcohol (**44**)

810 mg (33.77 mmol) NaH are suspended in 90 mL dry THF at room temperature, and 4.2 g (30.39 mmol) solid 1,4-bis(hydroxymethyl)-benzene is added in three portions over 5 min, and the reaction mixture is stirred for 45 min. 4.83 g (32.08 mmol) tert-butyldimethylsilyl chloride are added in three portions over 5 min and stirred for an additional 1.5 h before the mixture is quenched with water and then diluted with 100 mL of water and 100 mL of diethyl ether. The organic phase is separated and the aqueous phase is extracted with diethyl ether. The combined organic phases are washed with brine, dried over $MgSO_4$, filtered and concentrated *in vacuo*. The product is purified by flash chromatography (petrol ether/ethyl acetate 10:1) to yield 3.0 g **44** (11.88 mmol, 40%).

Example 12: 1-[(tert-Butyldimethylsilyloxy)methyl]-4-(iodomethyl)benzene (**45**)

9.15 g (34.88 mmol) triphenylphosphine and 3.2 g (44.5 mmol) imidazole are dissolved in a 3:1 mixture of diethyl ether/acetonitrile (30 mL). 8.85 g (34.9 mmol) iodine are added under vigorous stirring until a yellow suspension has formed. A solution of 6.1 g (23.25 mmol) of the monoprotected benzyl alcohol **44** in 20 mL of the same solvent mixture is added, and the mixture is stirred for 2 h. The solid is removed by filtration, the filtrate diluted with 100 mL of diethyl ether and washed with 100 mL of a saturated solution of sodium bisulfite. The aqueous solution is back-extracted with diethyl ether, the combined organic phases dried over $MgSO_4$ and concentrated *in vacuo*. Flash chromatography (petrol ether/ethyl acetate 95:5) yields 4.8 g **45** (13.25 mmol, 57%).

Example 13: 4-(13-Azido-2,5,8,11-tetraoxatridecyl)-benzyl alcohol (**46**)

4.8 g (13.25 mmol) 1-[(tert-butyldimethylsilyloxy)methyl]-4-(iodomethyl)benzene is dissolved in 70 mL dry THF under argon, and 0.954 g (39.75 mmol) NaH is added in small

portions over 10 min. A solution of 3.2 g (14.58 mmol) 1-azido-11-hydroxy-3,6,9-trioxaundecane (**7**) in 20 mL dry THF is added dropwise, and the reaction mixture stirred for 15 h at room temperature. 2 mL of water are added to quench the reaction and the mixture is concentrated to about 50% under reduced pressure. 70 mL of water are added and extracted with diethyl ether. The organic phase is dried over $MgSO_4$ and the solvent removed. Purification by silica gel chromatography (gradient: petrol ether/ethyl acetate 10:1 → 3:1) yields 3.8 g (8.38 mmol, 63%) of the TBDMS-protected product. It is dissolved in 80 mL dry THF in a plastic tube, and cooled to 0°C, and 8 mL of a pyridine/HF (70:30) solution is added and stirred for 3 h at room temperature. 100 mL of aqueous saturated $NaHCO_3$ are added, the organic phase separated, washed with brine and dried over $MgSO_4$. After removal of the solvent the product is purified by silica gel chromatography (petrol ether/ethyl acetate 1:1) to yield 1.27 g **46** (2.87 mmol, 74%).

Example 14: O^6-[4-(13-azido-2,5,8,11-tetraoxatridecyl)-oxymethyl-benzyl]guanine (**41**)

0.974 g (2.87 mmol) 4-(13-azido-2,5,8,11-tetraoxa-tridecyl)-benzyl alcohol is dissolved in 5 mL dry DMF and 1.3 g (11.5 mmol) potassium tert-butoxide are added. 0.731 g (2.87 mmol) of 1-(2-amino-7*H*-purin-6-yl)-1-methyl-pyrrolidinium chloride are then added and the solution stirred for 22 h. After removal of the solvent *in vacuo* the crude product is purified by flash column chromatography (methanol/dichloromethane 5:95). Yield: 0.675 g (50%).

Example 15: Hetero-benzylguanine-dimer **42**

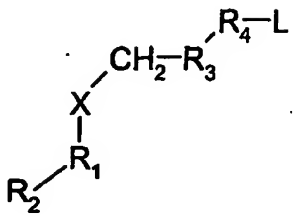
To a solution of 45 mg (0.09 mmol) azide **41** and 29.5 mg (0.09 mmol) O^6-(4-prop-2-ynyloxymethyl-benzyl)guanine (**35**) in 0.8 mL DMF is added a suspension of 9.8 mg CuI in 0.1 mL water, and the reaction mixture is stirred for 24 h at room temperature.

Example 16: Homo-benzylguanine-dimer **43**

To a solution of 50 mg (0.11 mmol) azide **41** and 5.6 mg (0.026 mmol) 1,4-diprop-2-ynyloxymethylbenzene (**40**) in 0.8 mL DMF is added a suspension of 3 mg CuI in 0.1 mL water and the reaction mixture is stirred at 40° C for 24 h.

Claims

1. A compound of the formula 1



1

wherein R_1-R_2 is a group recognized by AGT as a substrate;
X is oxygen or sulfur;
R_3 is a an aromatic or a heteroaromatic group, or an optionally substituted unsaturated alkyl, cycloalkyl or heterocyclyl group with the double bond connected to CH_2;
R_4 is a linker; and
L is a label, a bond connecting R_4 to R_1 forming a cyclic substrate, or a further group $-R_3-CH_2-X-R_1-R_2$.

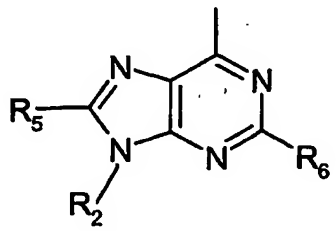
2. A compound of the formula 1 according to claim 1, wherein
R_1 is a heteroaromatic group containing 1 to 5 nitrogen atoms;
R_2 is hydrogen, alkyl of 1 to 10 carbon atoms, or a saccharide moiety;
X is oxygen;
R_3 is phenyl, an unsubstituted or substituted mono- or bicyclic heteroaryl group of 5 or 6 rings atoms comprising zero, one, two, three or four ring nitrogen atoms and zero or one oxygen atom and zero or one sulfur atom, with the proviso that at least one ring carbon atom is replaced by a nitrogen, oxygen or sulfur atom, 1-alkenyl, 1-alkinyl, 1-cyclohexenyl with 3 to 7 carbon atoms, or an optionally substituted unsaturated heterocyclyl group with 3 to 12 atoms and 1 to 5 heteroatoms selected from nitrogen, oxygen and sulfur, and a double bond in the position connecting the heterocyclyl group to methylene CH_2;

R_4 is an optionally substitued straight or branched chain alkylene group with 1 to 300 carbon atoms, wherein optionally
(a) one or more carbon atoms are replaced by oxygen
(b) one or more carbon atoms are replaced by nitrogen carrying a hydrogen atom, and the adjacent carbon atoms are substituted by oxo, representing an amide function –NH-CO-;

(c) one or more carbon atoms are replaced by oxygen, and the adjacent carbon atoms are substituted by oxo, representing an ester function –O-CO-;
(d) the bond between two adjacent carbon atoms is a double or a triple bond, representing a function –CH=CH- or –C≡C-;
(e) one or more carbon atoms are replaced by a phenylene, a saturated or unsaturated cycloalkylene, a saturated or unsaturated bicycloakylene, a bridging heteraromatic or a bridging saturated or unsaturated heterocyclyl group; and/or
(f) two adjacent carbon atoms are replaced by a disulfide linkage -S-S-; and

L is a spectroscopic probe, a magnetic probe, a contrast reagent, a radioactively labelled molecule, a molecule which is one part of a specific binding pair which is capable of specifically binding to a partner, a molecule that is suspected to interact with other biomolecules, a library of molecules that are suspected to interact with other biomolecules, a molecule which is capable of crosslinking to other molecules, a molecule which is capable of generating hydroxyl radicals upon exposure to H_2O_2 and ascorbate, a molecule which is capable of generating reactive radicals upon irradiation with light, a molecule covalently attached to a solid support, a nucleic acid or a derivative thereof capable of undergoing base-pairing with its complementary strand, a lipid or other hydrophobic molecule with membrane-inserting properties, or a biomolecule with enzymatic properties, a bond connecting R_4 to R_1 forming a cyclic substrate, or a further group $-R_3-CH_2-X-R_1-R_2$.

3. A compound of the formula 1 according to claim 1, wherein
R_1-R_2 is a radical of the formula 2



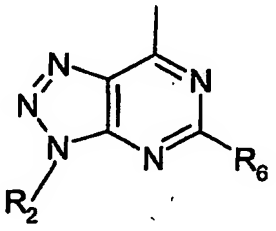
2

wherein R_2 is hydrogen, alkyl of 1 to 10 carbon atoms, or a saccharide moiety;
R_5 is hydrogen, halogen, trifluoromethyl, or hydroxy; and
R_6 is hydrogen, hydroxy or unsubstituted or substituted amino;
and tautomeric forms thereof.

4. A compound of formula 1 according to claim 3, wherein the saccharide moiety R_2 is a β-D-2'-deoxyribosyl, or a β-D-2'-deoxyribosyl being incorporated into a single stranded

oligodeoxyribonucleotide having a length of 2 to 99 nucleotides, wherein the guanine derivative R_1 occupies any position within the oligonucleotide sequence.

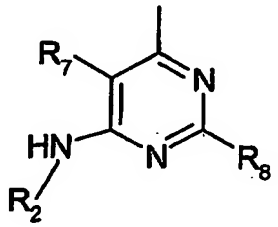
5. A compound of the formula 1 according to claim 1, wherein R_1-R_2 is a radical of the formula 3



3

wherein R_2 is hydrogen, alkyl of 1 to 10 carbon atoms, or a saccharide moiety; and R_6 is hydroxy or unsubstituted or substituted amino; and tautomeric forms thereof.

6. A compound of the formula 1 according to claim 1, wherein R_1-R_2 is a radical of the formula 4



4

wherein R_2 is hydrogen, alkyl of 1 to 10 carbon atoms, or a saccharide moiety; and R_7 and R_8 are both independently of one another hydrogen, halogen, lower alkyl with 1 to 4 carbon atoms, amino, or nitro.

7. A compound of the formula 1 according to claim 1 wherein R_3 is triazolyl, tetrazolyl, isoxazolyl, thienyl, or isoxazolidinyl.

8. A compound of the formula 1 according to claim 1 wherein R_4 is a straight chain alkylene group with 2 to 25 carbon atoms, a straight chain polyethylene glycol group with 4 to 100 ethyleneoxy units, or a straight chain alkylene group with 2 to 25 carbon atoms wherein two or more carbon atoms are replaced by an amide function –NH-CO, optionally attached to the group R_3 by a –CH=CH- or –C≡C- group.

9. A compound of the formula 1 according to claim 3 wherein R_3 is phenylene and L is a further group $-R_3-CH_2-X-R_1-R_2$.

10. A compound of the formula 1 according to claim 4 wherein R_3 is phenylene, R_6 is amino and L is a bond connecting R_4 to R_6.

11. A method for detecting and manipulating a protein of interest, characterized in that the protein of interest incorporated into a AGT fusion protein is contacted with an AGT substrates carrying a label, and the AGT fusion protein is detected and optionally further manipulated using the label in a system designed for recognising or handling the label, and wherein the AGT substrate carrying the label is a O^6-substituted guanine derivative or related nitrogen containing substituted hydroxy-heterocycle or a sulfur analog, and the O^6-substitutent is an activated methyl derivative suitable for transfer from guanine or the corresponding heterocycle to AGT.

12. A method according to claim 11 wherein the AGT substrate carrying a label is a compound of formula 1 according to claim 1.

Abstract

The present invention relates to methods of transferring a label from novel substrates to O^6-alkylguanine-DNA alkyltransferases (AGT) and O^6-alkylguanine-DNA alkyltransferase fusion proteins, and to novel substrates suitable in such methods. Proteins of interest are incorporated into a AGT fusion protein, the AGT fusion protein is contacted with particular AGT substrates carrying a label, and the AGT fusion protein is detected and optionally further manipulated using the label in a system designed for recognising and/or handling the label. The particular AGT substrates used in the method of the invention are O^6-substituted guanine derivatives or related nitrogen containing hydroxy-heterocycles and their sulfur analogs wherein the O^6-substitutent is an activated methyl derivative suitable for transfer from guanine or the corresponding heterocycle to AGT, and further carrying a label. The invention relates also to the novel AGT substrates as such, to methods of manufacture of such novel substrates, and to intermediates useful in the synthesis of such novel AGT substrates.